

#4



ENTERED

See page 6

PCT09

RAW SEQUENCE LISTING

DATE: 01/27/2003

PATENT APPLICATION: US/09/786,880C

TIME: 13:20:59

Input Set : A:\446.001.txt

Output Set: N:\CRF4\01272003\I786880C.raw

```

3 <110> APPLICANT: AVENTIS PHARMA S.A.
5 <120> TITLE OF INVENTION: ESSENTIAL GENES FROM C. ALBICANS AND A METHOD FOR
6   SCREENING ANTIMYCOTIC SUBSTANCES USING SAID GENES
8 <130> FILE REFERENCE: 446.001
10 <140> CURRENT APPLICATION NUMBER: 09/786,880C
11 <141> CURRENT FILING DATE: 2001-03-08
13 <150> PRIOR APPLICATION NUMBER: EP98402255.8
14 <151> PRIOR FILING DATE: 1998-09-11
16 <160> NUMBER OF SEQ ID NOS: 25
18 <170> SOFTWARE: PatentIn Ver. 2.1
20 <210> SEQ ID NO: 1
21 <211> LENGTH: 399
22 <212> TYPE: DNA
23 <213> ORGANISM: Artificial Sequence
25 <220> FEATURE:
26 <223> OTHER INFORMATION: Description of Artificial Sequence: Probe
28 <400> SEQUENCE: 1
29 attcatccca tcagtcgaga aagtttgcac agccatttac aacaattaat aaatgataaa 60
31 cctcaagaga cagtacaaga atcgtctgat ttattacaat ttatcccagt ctctagatta 120
33 cctgtcaaag ataataatgtt gaaatttgat caaattaatc ataaatctcc tactttgatt 180
35 atgggtatat tgaatatgac tcctgattca tttagtgatg gtgggaaaca ttttgaaaaa 240
37 gaactagata atattgtgaa gcaggcagag aaattagtca gtgagggtgc tacgattatt 300
39 gacattggag gagtttccac acgaccagga agtggtgaac ccactgagga agaagaattg 360
41 gaacgtgtga ttccattaat tagagctatt cgtcaatca 399
44 <210> SEQ ID NO: 2
45 <211> LENGTH: 2367
46 <212> TYPE: DNA
47 <213> ORGANISM: Candida albicans
49 <220> FEATURE:
50 <221> NAME/KEY: CDS
51 <222> LOCATION: (1)..(2364)
53 <220> FEATURE:
54 <221> NAME/KEY: gene
55 <222> LOCATION: (1)..(2364)
56 <223> OTHER INFORMATION: gene CaNL256
58 <400> SEQUENCE: 2
59 atg ttg aaa aac gat acc gtt ttc act aaa gat att tct tgt acg gcg 48
60 Met Leu Lys Asn Asp Thr Val Phe Thr Lys Asp Ile Ser Cys Thr Ala
61 1 5 10 15
63 ata act ggt aaa gat gcc tgg aat cgg cca aca cca caa cca atc act 96
64 Ile Thr Gly Lys Asp Ala Trp Asn Arg Pro Thr Pro Gln Pro Ile Thr
65 20 25 30
67 ata tca tta tct ttc aat act gat ttc cat aag gca tcg gaa ttg gat 144

```

RAW SEQUENCE LISTING

DATE: 01/27/2003

PATENT APPLICATION: US/09/786,880C

TIME: 13:20:59

Input Set : A:\446.001.txt

Output Set: N:\CRF4\01272003\I786880C.raw

```

68 Ile Ser Leu Ser Phe Asn Thr Asp Phe His Lys Ala Ser Glu Leu Asp
69          35          40          45
71 aat ttg aaa tac tca att aat tat gct gtt att acc aga aat gta act 192
72 Asn Leu Lys Tyr Ser Ile Asn Tyr Ala Val Ile Thr Arg Asn Val Thr
73          50          55          60
75 gaa ttt atg aaa tca aat gag cat tta aat ttc aag tca tta gga aat 240
76 Glu Phe Met Lys Ser Asn Glu His Leu Asn Phe Lys Ser Leu Gly Asn
77 65          70          75          80
79 att gct caa gca att agt gat att gga tta gat caa tct aga ggt ggt 288
80 Ile Ala Gln Ala Ile Ser Asp Ile Gly Leu Asp Gln Ser Arg Gly Gly
81          85          90          95
83 gga tct att gtg gat gtg acg ata aaa agt ttg aaa tca gaa ata aga 336
84 Gly Ser Ile Val Asp Val Thr Ile Lys Ser Leu Lys Ser Glu Ile Arg
85          100          105          110
87 gct gaa agt gtc gaa tat aaa att aat aga aac act ttg ggt caa ccc 384
88 Ala Glu Ser Val Glu Tyr Lys Ile Asn Arg Asn Thr Leu Gly Gln Pro
89          115          120          125
91 gtt cca tta gat att ttc caa gtt aat aaa ttg aga tta ttg acg att 432
92 Val Pro Leu Asp Ile Phe Gln Val Asn Lys Leu Arg Leu Leu Thr Ile
93          130          135          140
95 att gga gtt ttc aca ttt gaa aga tta caa aaa caa ata gtt gat gtt 480
96 Ile Gly Val Phe Thr Phe Glu Arg Leu Gln Lys Gln Ile Val Asp Val
97 145          150          155          160
99 gat ttg caa ttt aaa att gaa cct aat tcc aat tta tat ttc cat caa 528
100 Asp Leu Gln Phe Lys Ile Glu Pro Asn Ser Asn Leu Tyr Phe His Gln
101          165          170          175
103 ata att gct gat att gtt tca tac gtg gaa tca tct aat ttc aaa act 576
104 Ile Ile Ala Asp Ile Val Ser Tyr Val Glu Ser Ser Asn Phe Lys Thr
105          180          185          190
107 gta gaa gca ttg gtg tct aag att ggt caa ttg aca ttt cag aaa tat 624
108 Val Glu Ala Leu Val Ser Lys Ile Gly Gln Leu Thr Phe Gln Lys Tyr
109 195          200          205
111 gac gga gta gct gaa gtt gtt gct act gtc act aaa ccg aat gca ttt 672
112 Asp Gly Val Ala Glu Val Val Ala Thr Val Thr Lys Pro Asn Ala Phe
113          210          215          220
115 agt cat gtt gaa ggt gtt gga gta tca tct acc atg gtc aaa gac aat 720
116 Ser His Val Glu Gly Val Gly Val Ser Ser Thr Met Val Lys Asp Asn
117 225          230          235          240
119 ttc aaa gat atg gaa cca gtt aaa ttt gaa aac aca att gct caa act 768
120 Phe Lys Asp Met Glu Pro Val Lys Phe Glu Asn Thr Ile Ala Gln Thr
121          245          250          255
123 aat aga gca ttc aat tta cct gtt gaa aat gag aaa act gag gat tat 816
124 Asn Arg Ala Phe Asn Leu Pro Val Glu Asn Glu Lys Thr Glu Asp Tyr
125          260          265          270
127 acc ggg tac cac act gca ttt att gcc ttt gga tcc aat act gga aat 864
128 Thr Gly Tyr His Thr Ala Phe Ile Ala Phe Gly Ser Asn Thr Gly Asn
129          275          280          285
131 caa gta gaa aat att acc aat tca ttc gaa ttg ttg caa aaa tat gga 912
132 Gln Val Glu Asn Ile Thr Asn Ser Phe Glu Leu Leu Gln Lys Tyr Gly

```

RAW SEQUENCE LISTING

DATE: 01/27/2003

PATENT APPLICATION: US/09/786,880C

TIME: 13:20:59

Input Set : A:\446.001.txt

Output Set: N:\CRF4\01272003\I786880C.raw

133	290	295	300	
135	atc acc ata gaa gca act tca tca ttg tac att tct aaa cca atg tat	960		
136	Ile Thr Ile Glu Ala Thr Ser Ser Leu Tyr Ile Ser Lys Pro Met Tyr			
137	305	310	315	320
139	tac ttg gat caa cca gat ttt ttc aat gga gta att aaa gtg aat ttc	1008		
140	Tyr Leu Asp Gln Pro Asp Phe Phe Asn Gly Val Ile Lys Val Asn Phe			
141		325	330	335
143	caa aac att tca cct ttc cag ttg ttg aaa att cta aaa gat att gaa	1056		
144	Gln Asn Ile Ser Pro Phe Gln Leu Leu Lys Ile Leu Lys Asp Ile Glu			
145		340	345	350
147	tat aaa cat tta gaa agg aaa aaa gac ttt gat aat ggg ccc aga tca	1104		
148	Tyr Lys His Leu Glu Arg Lys Lys Asp Phe Asp Asn Gly Pro Arg Ser			
149		355	360	365
151	ata gat ttg gat att ata cta tat gac gat tta caa tta aat acc gag	1152		
152	Ile Asp Leu Asp Ile Ile Leu Tyr Asp Asp Leu Gln Leu Asn Thr Glu			
153		370	375	380
155	aat cta att att cca cat aaa tca atg tta gaa aga aca ttt gta tta	1200		
156	Asn Leu Ile Ile Pro His Lys Ser Met Leu Glu Arg Thr Phe Val Leu			
157	385	390	395	400
159	caa cca tta tgt gaa gta ttg ccc cct gat tat att cat ccc atc agt	1248		
160	Gln Pro Leu Cys Glu Val Leu Pro Pro Asp Tyr Ile His Pro Ile Ser			
161		405	410	415
163	gca gaa agt ttg cat agc cat tta caa caa tta ata aat gat aaa cct	1296		
164	Ala Glu Ser Leu His Ser His Leu Gln Gln Leu Ile Asn Asp Lys Pro			
165		420	425	430
167	caa gag aca gta caa gaa tcg tct gat tta tta caa ttt atc cca gtc	1344		
168	Gln Glu Thr Val Gln Glu Ser Ser Asp Leu Leu Gln Phe Ile Pro Val			
169		435	440	445
171	tct aga ttg cct gtc aaa gat aat att ttg aaa ttt gat caa att aat	1392		
172	Ser Arg Leu Pro Val Lys Asp Asn Ile Leu Lys Phe Asp Gln Ile Asn			
173		450	455	460
175	cat aaa tct cct act ttg att atg ggt ata ttg aat atg act cct gat	1440		
176	His Lys Ser Pro Thr Leu Ile Met Gly Ile Leu Asn Met Thr Pro Asp			
177	465	470	475	480
179	tca ttt agt gat ggt ggg aaa cat ttt gga aaa gaa cta gat aat act	1488		
180	Ser Phe Ser Asp Gly Gly Lys His Phe Gly Lys Glu Leu Asp Asn Thr			
181		485	490	495
183	gtg aag cag gca gag aaa tta gtc agt gag ggt gct acg att att gac	1536		
184	Val Lys Gln Ala Glu Lys Leu Val Ser Glu Gly Ala Thr Ile Ile Asp			
185		500	505	510
187	att gga gga gtt tcc aca cgc cca gga agt gtt gaa ccc act gag gaa	1584		
188	Ile Gly Gly Val Ser Thr Arg Pro Gly Ser Val Glu Pro Thr Glu Glu			
189		515	520	525
191	gaa gaa ttg gaa cgt gtg att cca tta att aaa gct att cgt caa tca	1632		
192	Glu Glu Leu Glu Arg Val Ile Pro Leu Ile Lys Ala Ile Arg Gln Ser			
193		530	535	540
195	ctg aac cct gat tta ctg aag gtg ttg att tcg gtt gat act tat cgt	1680		
196	Leu Asn Pro Asp Leu Leu Lys Val Leu Ile Ser Val Asp Thr Tyr Arg			
197	545	550	555	560

RAW SEQUENCE LISTING

DATE: 01/27/2003

PATENT APPLICATION: US/09/786,880C

TIME: 13:20:59

Input Set : A:\446.001.txt

Output Set: N:\CRF4\01272003\I786880C.raw

```

199 agg aac gtt gct gaa caa agt tta ctt gtg ggt gct gac ata atc aac 1728
200 Arg Asn Val Ala Glu Gln Ser Leu Leu Val Gly Ala Asp Ile Ile Asn
201          565          570          575
203 gat atc tca atg ggc aaa tat gat gaa aaa ata ttt gat gtg gtt gct 1776
204 Asp Ile Ser Met Gly Lys Tyr Asp Glu Lys Ile Phe Asp Val Val Ala
205          580          585          590
207 aaa tac gga tgt cct tat atc atg aat cat act cga gga tca cct aaa 1824
208 Lys Tyr Gly Cys Pro Tyr Ile Met Asn His Thr Arg Gly Ser Pro Lys
209          595          600          605
211 acc atg tct aaa ttg acc aat tat gaa tca aat aca aat gat gat att 1872
212 Thr Met Ser Lys Leu Thr Asn Tyr Glu Ser Asn Thr Asn Asp Asp Ile
213          610          615          620
215 atc gaa tat ata att gat cct aaa tta gga cat caa gaa ttg gat ttg 1920
216 Ile Glu Tyr Ile Ile Asp Pro Lys Leu Gly His Gln Glu Leu Asp Leu
217 625          630          635          640
219 tca cct gaa atc aag aat tta ctc aat gga atc agt cgt gaa ttg agt 1968
220 Ser Pro Glu Ile Lys Asn Leu Leu Asn Gly Ile Ser Arg Glu Leu Ser
221          645          650          655
223 tta caa atg ttt aaa gcc atg gct aaa gga gtg aaa aaa tgg caa att 2016
224 Leu Gln Met Phe Lys Ala Met Ala Lys Gly Val Lys Lys Trp Gln Ile
225          660          665          670
227 att ttg gat cct ggt att gga ttt gct aaa aat ttg aat caa aat tta 2064
228 Ile Leu Asp Pro Gly Ile Gly Phe Ala Lys Asn Leu Asn Gln Asn Leu
229          675          680          685
231 gca gtt att cgt aat gcc tcg ttt ttt aaa aaa tat tct att caa att 2112
232 Ala Val Ile Arg Asn Ala Ser Phe Phe Lys Lys Tyr Ser Ile Gln Ile
233          690          695          700
235 aat gaa cgt gtt gat gat gtg aca atc aaa cat aaa tat tta agt ttt 2160
236 Asn Glu Arg Val Asp Asp Val Thr Ile Lys His Lys Tyr Leu Ser Phe
237 705          710          715          720
239 aat ggt gct tgt gtt ttg gtg ggg aca tca aga aag aag ttt ttg ggg 2208
240 Asn Gly Ala Cys Val Leu Val Gly Thr Ser Arg Lys Lys Phe Leu Gly
241          725          730          735
243 aca tta act ggt aat gaa gtg cct ctg gat cga gta ttt ggc act ggt 2256
244 Thr Leu Thr Gly Asn Glu Val Pro Leu Asp Arg Val Phe Gly Thr Gly
245          740          745          750
247 gca aca gtg tct gcg tgt att gaa caa aac act gat att gta aga gtt 2304
248 Ala Thr Val Ser Ala Cys Ile Glu Gln Asn Thr Asp Ile Val Arg Val
249          755          760          765
251 cat gat gtt aaa gaa atg aaa gat gta gta tgt ata agt gat gca att 2352
252 His Asp Val Lys Glu Met Lys Asp Val Val Cys Ile Ser Asp Ala Ile
253          770          775          780
255 tat aaa aat gta taa 2367
256 Tyr Lys Asn Val
257 785
260 <210> SEQ ID NO: 3
261 <211> LENGTH: 788
262 <212> TYPE: PRT
263 <213> ORGANISM: Candida albicans

```

RAW SEQUENCE LISTING

DATE: 01/27/2003

PATENT APPLICATION: US/09/786,880C

TIME: 13:20:59

Input Set : A:\446.001.txt

Output Set: N:\CRF4\01272003\I786880C.raw

```

265 <400> SEQUENCE: 3
266 Met Leu Lys Asn Asp Thr Val Phe Thr Lys Asp Ile Ser Cys Thr Ala
267   1           5           10           15
269 Ile Thr Gly Lys Asp Ala Trp Asn Arg Pro Thr Pro Gln Pro Ile Thr
270           20           25           30
272 Ile Ser Leu Ser Phe Asn Thr Asp Phe His Lys Ala Ser Glu Leu Asp
273           35           40           45
275 Asn Leu Lys Tyr Ser Ile Asn Tyr Ala Val Ile Thr Arg Asn Val Thr
276           50           55           60
278 Glu Phe Met Lys Ser Asn Glu His Leu Asn Phe Lys Ser Leu Gly Asn
279           65           70           75           80
281 Ile Ala Gln Ala Ile Ser Asp Ile Gly Leu Asp Gln Ser Arg Gly Gly
282           85           90           95
284 Gly Ser Ile Val Asp Val Thr Ile Lys Ser Leu Lys Ser Glu Ile Arg
285           100          105          110
287 Ala Glu Ser Val Glu Tyr Lys Ile Asn Arg Asn Thr Leu Gly Gln Pro
288           115          120          125
290 Val Pro Leu Asp Ile Phe Gln Val Asn Lys Leu Arg Leu Leu Thr Ile
291           130          135          140
293 Ile Gly Val Phe Thr Phe Glu Arg Leu Gln Lys Gln Ile Val Asp Val
294          145          150          155          160
296 Asp Leu Gln Phe Lys Ile Glu Pro Asn Ser Asn Leu Tyr Phe His Gln
297           165          170          175
299 Ile Ile Ala Asp Ile Val Ser Tyr Val Glu Ser Ser Asn Phe Lys Thr
300           180          185          190
302 Val Glu Ala Leu Val Ser Lys Ile Gly Gln Leu Thr Phe Gln Lys Tyr
303           195          200          205
305 Asp Gly Val Ala Glu Val Val Ala Thr Val Thr Lys Pro Asn Ala Phe
306           210          215          220
308 Ser His Val Glu Gly Val Gly Val Ser Ser Thr Met Val Lys Asp Asn
309          225          230          235          240
311 Phe Lys Asp Met Glu Pro Val Lys Phe Glu Asn Thr Ile Ala Gln Thr
312           245          250          255
314 Asn Arg Ala Phe Asn Leu Pro Val Glu Asn Glu Lys Thr Glu Asp Tyr
315           260          265          270
317 Thr Gly Tyr His Thr Ala Phe Ile Ala Phe Gly Ser Asn Thr Gly Asn
318           275          280          285
320 Gln Val Glu Asn Ile Thr Asn Ser Phe Glu Leu Leu Gln Lys Tyr Gly
321           290          295          300
323 Ile Thr Ile Glu Ala Thr Ser Ser Leu Tyr Ile Ser Lys Pro Met Tyr
324          305          310          315          320
326 Tyr Leu Asp Gln Pro Asp Phe Phe Asn Gly Val Ile Lys Val Asn Phe
327           325          330          335
329 Gln Asn Ile Ser Pro Phe Gln Leu Leu Lys Ile Leu Lys Asp Ile Glu
330           340          345          350
332 Tyr Lys His Leu Glu Arg Lys Lys Asp Phe Asp Asn Gly Pro Arg Ser
333           355          360          365
335 Ile Asp Leu Asp Ile Ile Leu Tyr Asp Asp Leu Gln Leu Asn Thr Glu
336           370          375          380

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/09/786,880C

DATE: 01/27/2003
TIME: 13:21:00

Input Set : A:\446.001.txt
Output Set: N:\CRF4\01272003\I786880C.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:12; N Pos. 338,339,340,341,342,344,347,349,366,368,383,387,407,408,433

VERIFICATION SUMMARY

DATE: 01/27/2003

PATENT APPLICATION: US/09/786,880C

TIME: 13:21:00

Input Set : A:\446.001.txt

Output Set: N:\CRF4\01272003\I786880C.raw

L:1944 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:300
L:1946 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:360
L:1948 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:420